

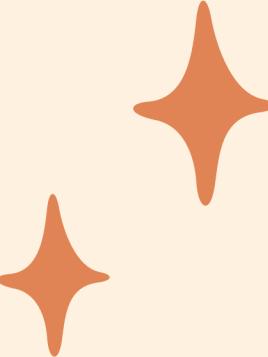
LINEAR REGRESSION

Do Ngoc Lam

NỘI DUNG

1. GIỚI THIỆU
2. BÀI TOÁN
3. ĐIỂM MẠNH
4. ĐIỂM YẾU
5. ỨNG DỤNG

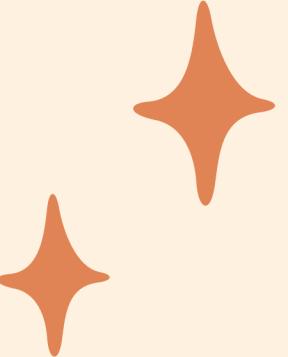




LINEAR REGRESSION

Là một phương pháp thống kê được sử dụng để mô hình hóa mối quan hệ giữa một biến phụ thuộc y vào một hoặc nhiều biến độc lập x



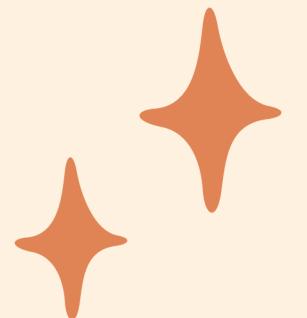


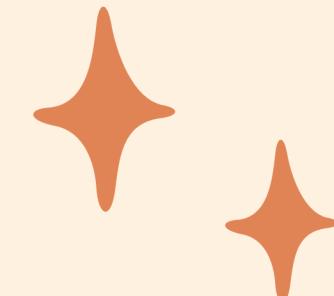
LINEAR REGRESSION

Mô hình được xây dựng dựa trên giả định
rằng y có thể biểu diễn dưới dạng một tổ
hợp tuyến tính của các biến x

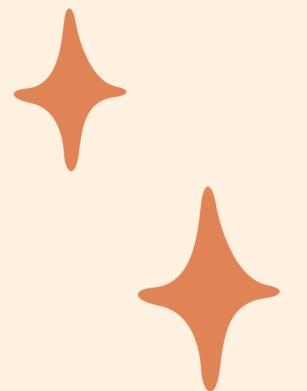
$$y = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n + \epsilon$$

BÀI TOÁN THỰC TẾ





MỐI QUAN HỆ CÂN NẶNG <--> CHIỀU CAO





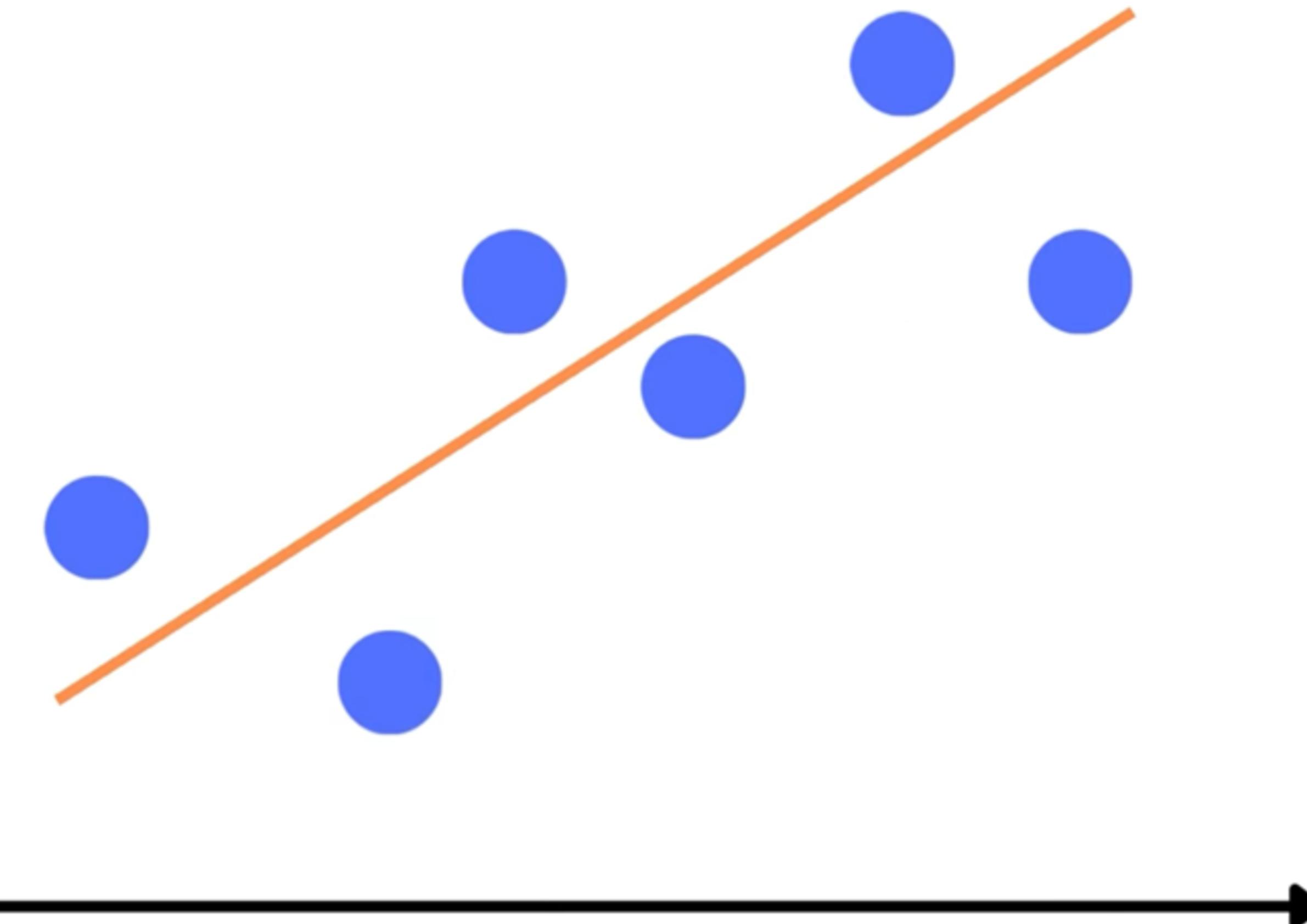
Càng cao,
Càng nặng

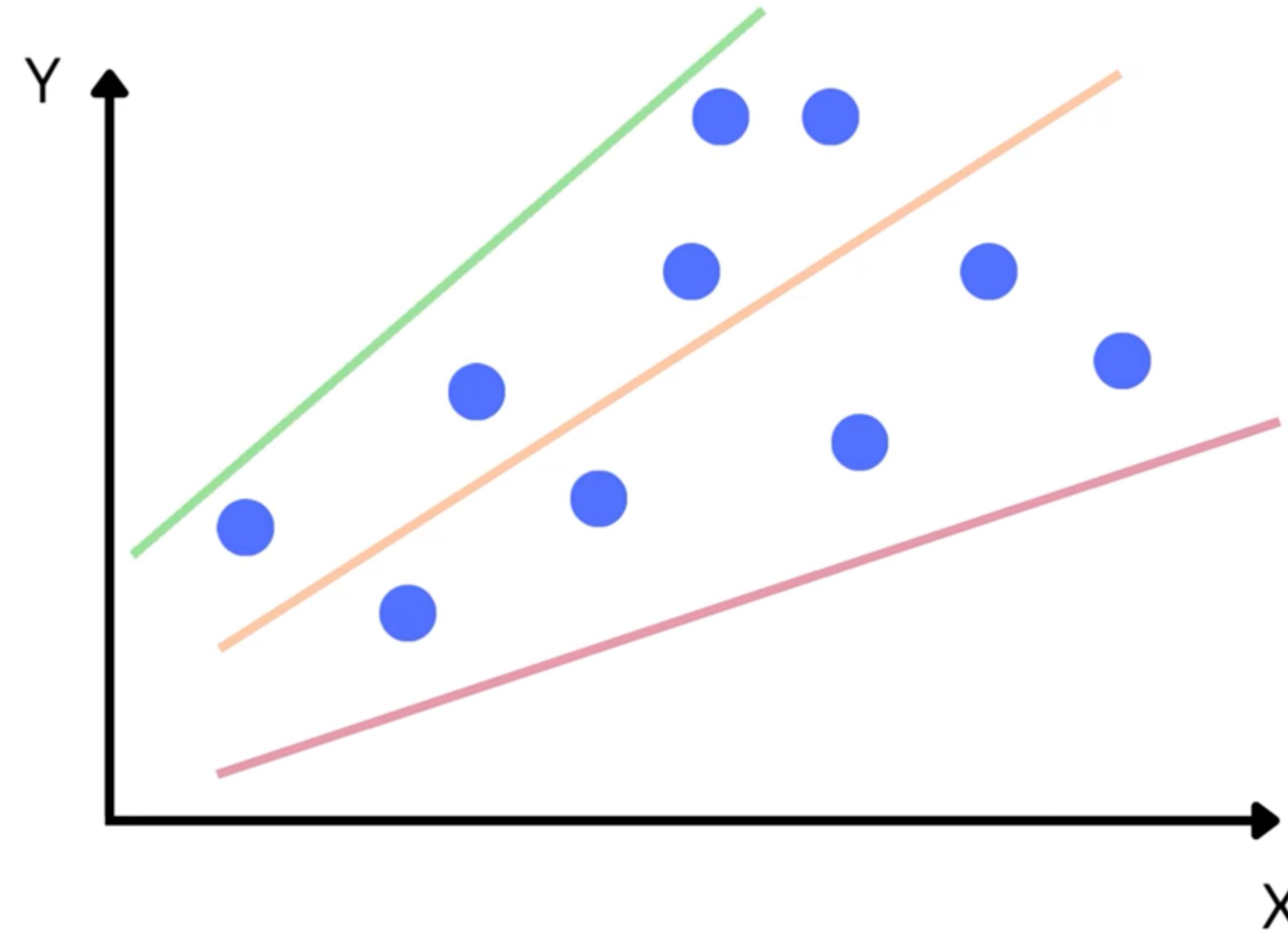
Chiều cao	Cân nặng
1.54m	45kg
1.6m	50km
1.78m	70km

Cân nặng



Chiều cao





“Tổng bình phương khoảng cách đến các điểm dữ liệu nhỏ nhất”

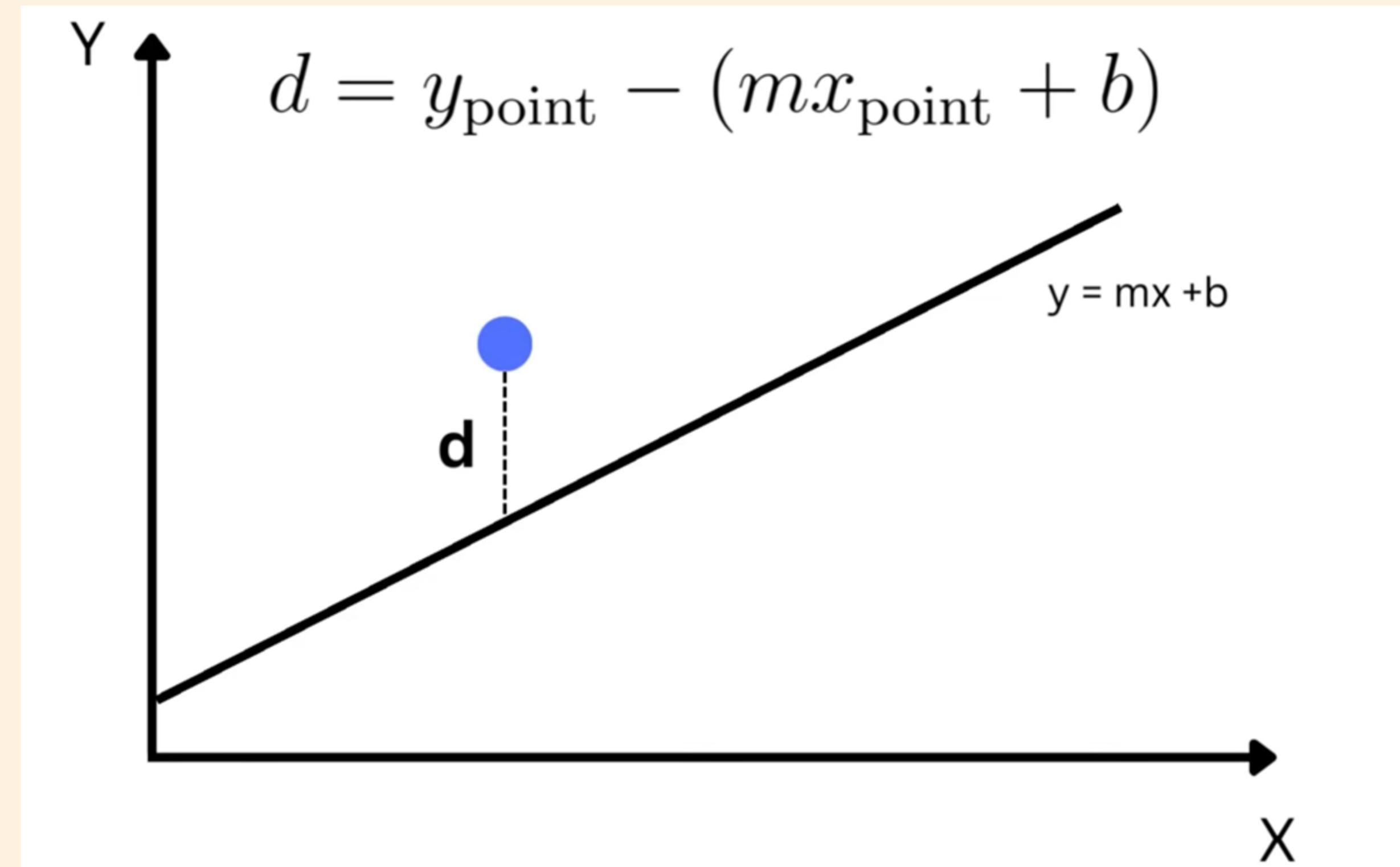
Y

$$d = y_{\text{point}} - (mx_{\text{point}} + b)$$

d

$$y = mx + b$$

X



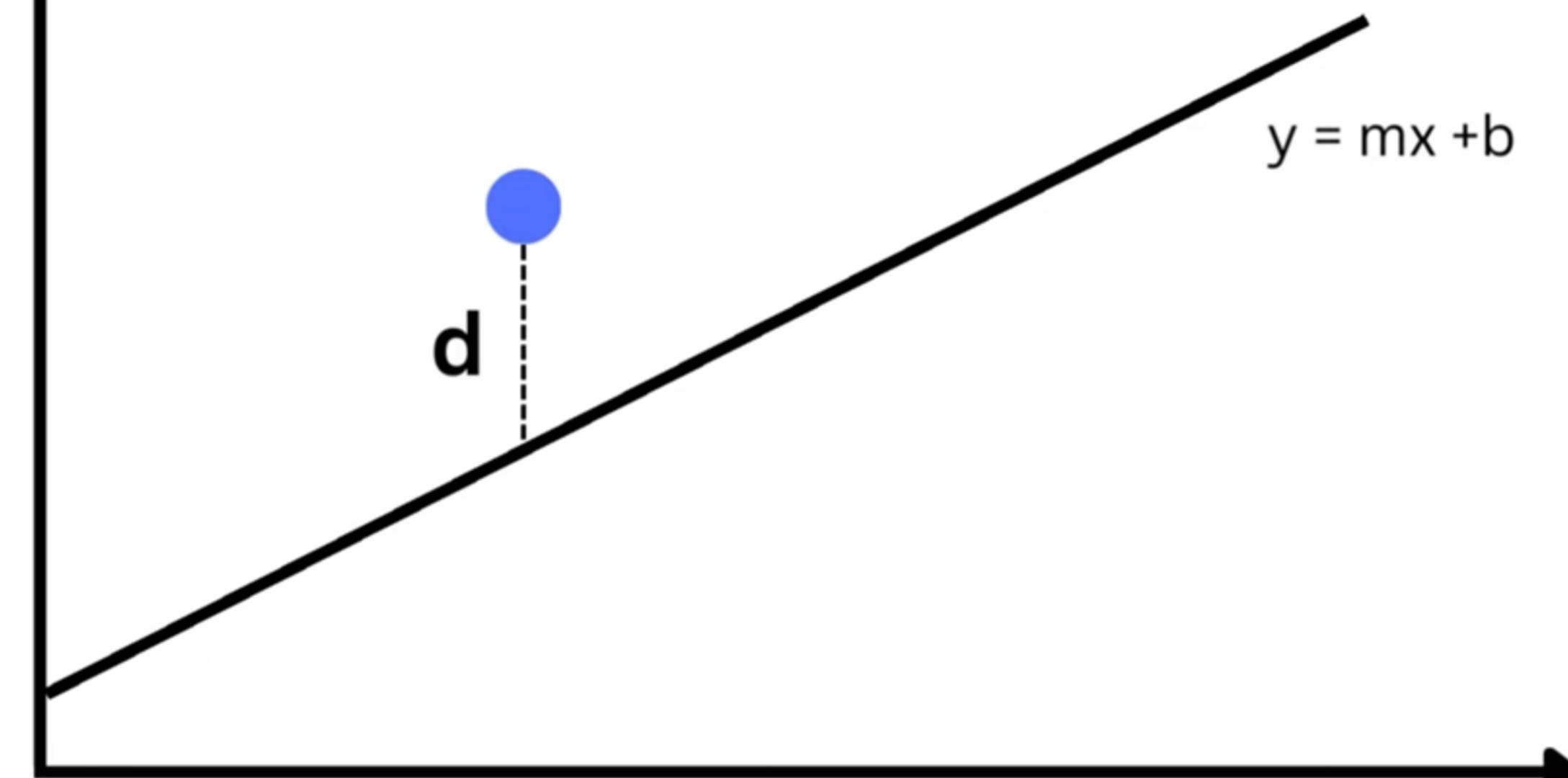
Y

$$d^2 = (y_{\text{point}} - (mx_{\text{point}} + b))^2$$

d

$$y = mx + b$$

X



$$Loss = \sum_{\text{all points}} (y_{\text{point}} - (mx_{\text{point}} + b))^2$$

DERIVATIVES

(ĐẠO HÀM)

**LOSS ĐẠT CỰC TIỂU
KHI ĐẠO HÀM TẠI
MÙA BẮNG 0**

$$\frac{\partial}{\partial m} \sum_{\text{all points}} (y_{\text{point}} - (mx_{\text{point}} + b))^2 = 0$$

$$\frac{\partial}{\partial b} \sum_{\text{all points}} (y_{\text{point}} - (mx_{\text{point}} + b))^2 = 0$$

$$m^* = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{N \sum x_i^2 - (\sum x_i)^2}$$

$$b^* = \frac{\sum y_i - m^* \sum x_i}{N}$$

N: Số điểm dữ liệu

x_i, y_i : Tọa độ mỗi điểm dữ liệu

DESCENT METHOD

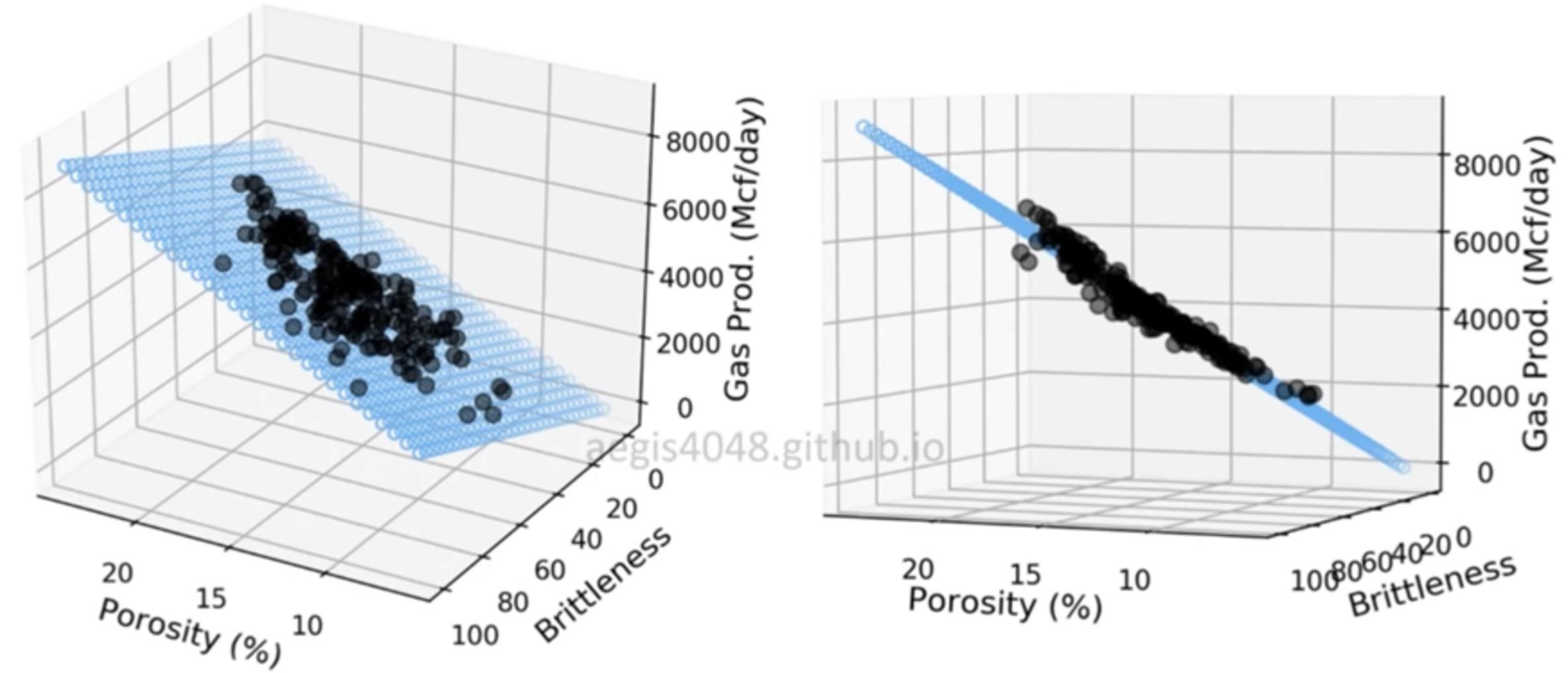
$$m^* = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{N \sum x_i^2 - (\sum x_i)^2}$$

$$b^* = \frac{\sum y_i - m^* \sum x_i}{N}$$

Lời giải trực tiếp
(analytical solution)

**TRONG THỰC TẾ, DESCENT
METHOD ĐƯỢC ƯA CHUỘNG
HƠN DO TỐN ÍT CÔNG SỨC
TÍNH TOÁN**

3D multiple linear regression model



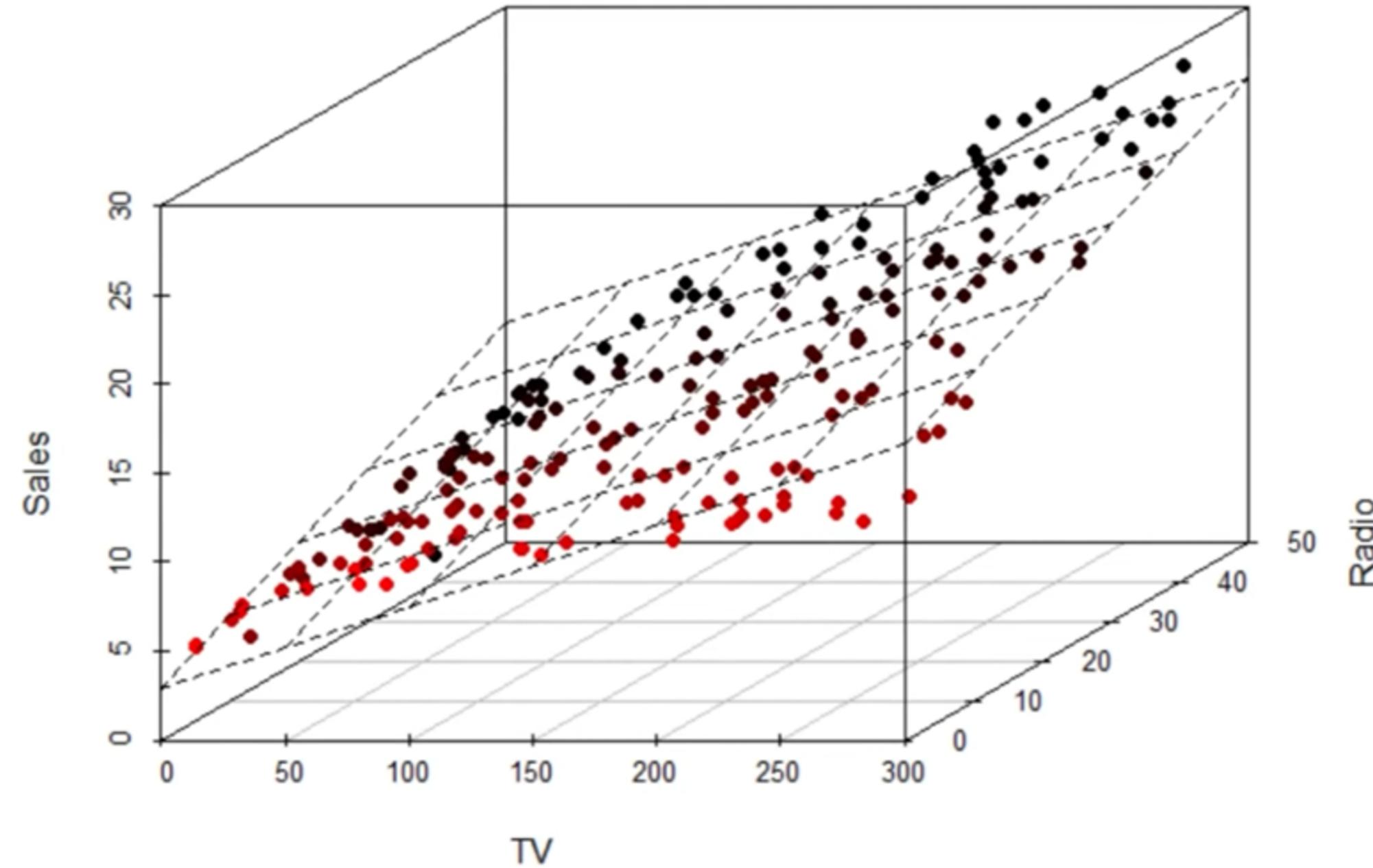
$$\text{minimize}_m \|Xm - y\|_2^2$$

$$y = m_1x_1 + m_2x_2 + b$$

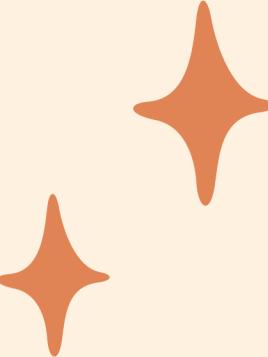
y: Giá của ngôi nhà

x₁: Diện tích

x₂: Số phòng ngủ



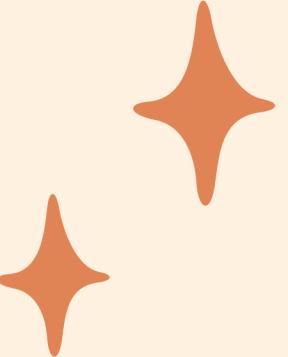
ĐIỂM MẠNH



ĐIỂM MẠNH

1. ĐƠN GIẢN, DỄ HIỂU
 2. HIỆU SUẤT CAO VỚI DỮ LIỆU NHỎ UÀ
TUYẾN TÍNH
 3. KHÔNG CẦN NHIỀU TÀI NGUYÊN TÍNH
TOÁN
 4. CÓ THỂ MỞ RỘNG SANG CÁC MÔ HÌNH
KHÁC
 5. TỐT ĐỂ LÀM BASELINE MODEL
- 
- 

ĐIỂM YÊU



ĐIỂM MẠNH

1. GIẢ ĐỊNH TUYỀN TÍNH KHÔNG PHẢI LÚC NÀO CŨNG ĐÚNG
 2. RẤT NHẠY CẢM UỚI NHIỄU
 3. KHÔNG HOẠT ĐỘNG TỐT UỚI DỮ LIỆU LỚN, PHỨC TẠP
 4. KHÔNG THỂ GIẢI QUYẾT UẨN ĐỀ PHÂN LOẠI
- 
- 

ỨNG DỤNG

ỨNG DỤNG

- DỰ ĐOÁN GIÁ TRỊ LIÊN TỤC**
- PHÂN TÍCH XU HƯỚNG**
- DỰ BÁO TÀI CHÍNH**
- KIỂM SOÁT CHẤT LƯỢNG**
- PHÂN TÍCH Y TẾ**



THANKIU AND HAUE FUN ~